

Universidad ORT Uruguay
Facultad de Ingeniería

Clusterización de Interacciones

Para la identificación de motivos de contacto

Entregado como requisito para la obtención del título de Máster en Big Data

Luciana Maschi – 165405

Tutores: Matías Sorozabal, Guzmán González

2023

Declaración de autoría

Yo, Luciana Maschi, declaro que el trabajo que se presenta en esa obra es de nuestra propia mano. Podemos asegurar que:

- La obra fue producida en su totalidad mientras realizaba el Proyecto Final del Máster en Big Data;
- Cuando he consultado el trabajo publicado por otros, lo he atribuido con claridad;
- Cuando he citado obras de otros, he indicado las fuentes. Con excepción de estas citas, la obra es enteramente mía;
- En la obra, he acusado recibo de las ayudas recibidas;
- Cuando la obra se basa en trabajo realizado conjuntamente con otros, he explicado claramente qué fue contribuido por otros, y qué fue contribuido por mí;
- Ninguna parte de este trabajo ha sido publicada previamente a su entrega, excepto donde se han realizado las aclaraciones correspondientes.



Luciana Maschi

11/09/2023

Agradecimientos

Quiero agradecer a todas las personas que contribuyeron de manera significativa a la realización de este proyecto.

A mis tutores, Matías Sorozabal y Guzmán González, por su constante apoyo y valiosas sugerencias que resultaron clave para llevar a cabo este proyecto. Su experiencia y compromiso fueron fundamentales.

A Telefónica Móviles del Uruguay, por proporcionar los datos y recursos necesarios. Su colaboración y disposición para compartir información fueron esenciales para lograr los objetivos.

A mis compañeros de trabajo, en particular al equipo de Business Intelligence, por su colaboración en el etiquetado de la fuente de datos. Su contribución fue de gran importancia.

A mi familia y amigos por la paciencia y comprensión a lo largo de este proceso. Sin su apoyo, no habría sido posible.

Abstract

Los datos se han convertido en un recurso invaluable para todas las organizaciones ya que la habilidad para recopilar, analizar y aprovechar datos de manera efectiva se ha vuelto esencial para la toma de decisiones y la optimización de procesos empresariales.

El presente trabajo se centra en la exploración técnicas de Procesamiento del Lenguaje Natural (NLP) con el propósito de identificar patrones recurrentes en las interacciones de los usuarios con el servicio de atención al cliente de una organización.

Luego de un preprocesamiento de los datos, donde se aplicaron diversas técnicas de limpieza y lematización, se desarrolló un modelo de clusterización. La construcción de este modelo involucró una secuencia de módulos, que abarcó desde la generación de *embeddings* utilizando Sentence-BERT, la reducción de dimensionalidad a través de UMAP, la clasificación mediante HDBSCAN, la caracterización y auto etiquetado utilizando C-TF-IDF y la posterior reducción de los clústers resultantes.

Se obtuvieron 34 clústers a partir de los cuales se han identificado oportunidades de mejora que permitirían generar eficiencias en la asignación de recursos, mejorar la experiencia del cliente y potenciar las ventas de productos y servicios.

Palabras clave

NLP, Lematización, Clusterización, Topic Modeling, Transformers, BERT, Sentence-BERT, Embeddings, UMAP, HDBSCAN, Grid Search, C-TF-IDF, Microsoft Azure, Azure Machine Learning.

Índice

1. Introducción	7
1.1. Objetivo	7
2. Infraestructura.....	9
2.1. Azure.....	9
3. Tratamiento de datos	11
3.1. Interacción	11
3.2. Análisis Exploratorio	11
3.3. Limpieza de Stopwords, caracteres especiales y mensajes automáticos	15
3.4. Lematización.....	16
3.5. Verbos	16
3.6. DataFrame de entrenamiento.	17
3.6.1. Descripción de las columnas.....	17
3.6.2. Descripción del contenido	17
4. Modelo	20
4.1. Transformers	20
4.2. BERT	21
4.3. Embeddings	21
4.4. Clasificación	Error! Bookmark not defined.
4.4.1. UMAP	22
4.4.2. HDBSCAN.....	22
4.5. Grid Search	22
4.6. Caracterización	24
4.6.1. C-TF-IDF	24
4.6.2. Auto etiquetado	25
4.7. Reducción de Tópicos	25
4.8. Resultados	25
5. Evaluación del Modelo	27
6. Conclusiones y próximos pasos	30
7. Bibliografía.....	Error! Bookmark not defined.
8. Anexos.....	33
8.1. Anexo 1: Casos de uso automáticos implementados.	33
8.2. Anexo 2: Ejemplo de Interacción.....	34
8.3. Anexo 3: Ejemplo Click To Chat.....	35
8.4. Anexo 4: Resultados Grid Search	36
8.5. Anexo 5: Matriz de similitud del coseno	37
8.6. Anexo 6: Interfaz de etiquetado.....	38
8.7. Anexo 7: Interacción con chatbot sobre activación de roaming	39

1. Introducción

Los datos se han transformado en un recurso de gran valor para todas las organizaciones. La capacidad de recopilar, analizar y utilizar datos de manera efectiva se ha vuelto fundamental para las empresas ya que les permite tomar decisiones más acertadas y optimizar sus procesos.

El presente trabajo se enmarca en una compañía de telefonía móvil, donde se explotarán los datos generados por las interacciones escritas de los clientes con el área de atención al cliente.

Diariamente se registra un promedio de 5500 interacciones por vías telefónicas, escritas o presenciales de clientes con atención al cliente y cada una de ellas genera información que es almacenada en una base de datos.

Para cada una de las interacciones escritas, además de almacenar datos sobre el cliente, asesor, fecha y hora de inicio y fin, se almacena cada uno de los mensajes intercambiados en la interacción, ya sean del cliente, asesor o chatbot.

Para los canales escritos existe una primera línea de atención donde un *chatbot* intenta resolver las consultas de los clientes con respuestas estandarizadas o mediante casos de uso automatizados y personalizados. Las consultas que el *chatbot* no logra resolver son derivadas a un asesor de atención al cliente. Mensualmente, un equipo compuesto por aproximadamente 100 asesores es responsable de responder estas consultas y brindar asistencia personalizada a los clientes.

Conocer los motivos de estas interacciones es fundamental para comprender y satisfacer las necesidades de los clientes, pero también ofrece la oportunidad de generar eficiencias en los procesos internos y mejorar la experiencia y satisfacción del cliente.

1.1. Objetivo

Hoy en día un 69% de los contactos son resueltos por un *chatbot* utilizando casos de uso automatizados personalizados y respuestas predefinidas para responder las consultas de los clientes. Es un objetivo de la organización lograr automatizar el 80% de las interacciones generando hasta 5 nuevos casos de uso. El listado de casos de uso existentes se puede consultar en el Anexo 1.

Para colaborar con la consecución del objetivo organizacional este proyecto tiene como objetivo desarrollar un modelo de identificación de motivos de contacto que sirva como prueba de concepto.

Se explotarán los datos almacenados de las interacciones escritas mediante técnicas de Procesamiento del Lenguaje Natural (NLP) y se buscará identificar motivos de contacto recurrentes que luego puedan ser utilizados para desarrollar nuevos casos de uso personalizados y así alcanzar los objetivos planteados por la organización.

Se utilizarán los contactos vía WhatsApp, que representan un 44% de las interacciones escritas, para desarrollar un modelo que sea capaz de identificar los principales motivos de contacto presentes.

2. Infraestructura

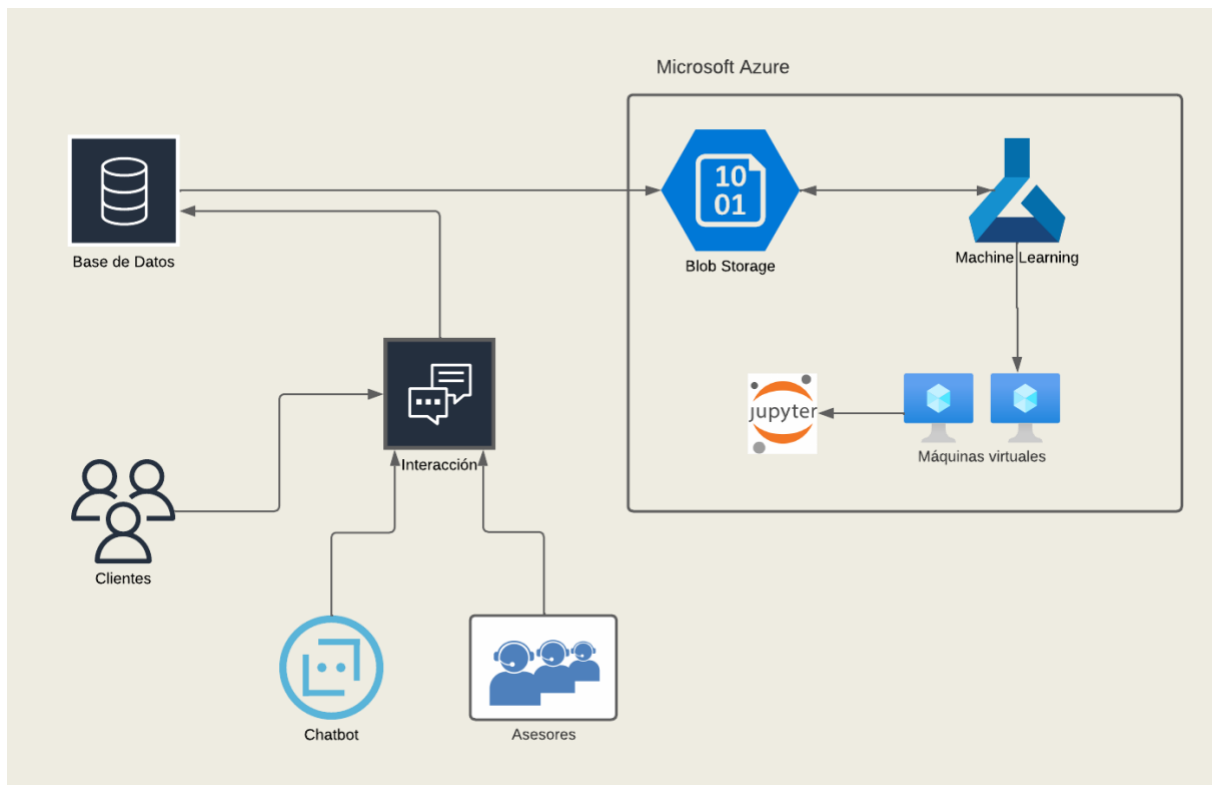


Ilustración 1. Diagrama de infraestructura

2.1. Azure

Azure es una plataforma en la nube de servicios de computación proporcionada por Microsoft. Para este proyecto se utiliza el servicio de Azure Machine Learning (Azure ML) y Azure Blob Storage.

Azure ML proporciona una plataforma completa para el desarrollo, entrenamiento, implementación y administración de modelos de aprendizaje automático e inteligencia artificial. Dentro de Azure ML se crearon dos máquinas virtuales, una para la preparación, desarrollo y entrenamiento del modelo y otra para exponer una interfaz de etiquetado manual.

Azure Blob Storage es un servicio de almacenamiento en la nube para almacenar datos no estructurados. Se utilizó para almacenar todos los elementos del proyecto, fuentes de datos, datos de entrenamiento, *embeddings*, modelos, etc. El acceso a dichos datos, está resguardado bajo un acuerdo de confidencialidad que debe firmar toda persona que requiera acceso a los mismos.

Azure es de principal ayuda en la fase de análisis de datos y en el desarrollo y entrenamiento de modelos ya que permite avanzar de manera rápida y eficiente. Dado que Azure es una herramienta que actualmente se utiliza en la organización una de las ventajas que ofrece es que si en el futuro se decide llevar el modelo a producción, el proceso sería bastante sencillo y escalable.

3. Tratamiento de datos

Los datos de las interacciones son almacenados en una tabla donde uno de los campos contiene un XML con el intercambio de mensajes de una interacción. En el Anexo 2 se puede visualizar la estructura de estos.

Inicialmente se obtuvo una extracción de datos correspondiente a los últimos 2 años. Sin embargo, con el propósito de optimizar el proceso de desarrollo y entrenamiento de los modelos, se decidió restringir el conjunto de datos a los últimos 6 meses. Se considera que un periodo de 6 meses proporciona un conjunto de datos representativo, dado que el objetivo es identificar los motivos de contacto más recurrentes, este período debería evidenciar con precisión dichos motivos en los datos.

3.1. Interacción

Se define una interacción como el intercambio de mensajes entre un cliente y un asesor de atención, ya sea humano o un *chatbot*.

Cada interacción está sujeta a un tiempo máximo de espera preestablecido: si no se reciben mensajes durante un período de un minuto, la interacción se da por finalizada de forma automática.

3.2. Análisis Exploratorio

Se utilizó la librería de Python ElementTree para poder extraer los datos de interés del XML. Por ejemplo, interesa poder distinguir los mensajes enviados por el cliente de los enviados por el agente, conocer si el agente involucrado fue el *chatbot* o un asesor humano y obtener la cantidad de mensajes intercambiados.

Se observa una tendencia creciente en la cantidad de interacciones, oscilando en torno a las 50.000 interacciones mensuales:

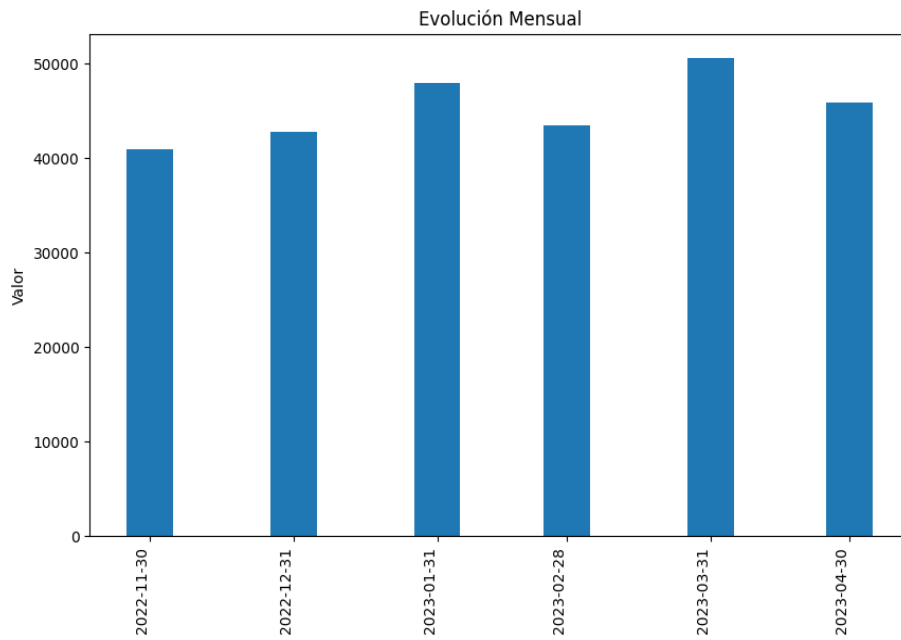


Ilustración 2. Evolución mensual de interacciones

En las interacciones vía WhatsApp el ratio de resolución del *chatbot* (57%) es menor al ratio relevado para el universo de interacciones escritas (69%):

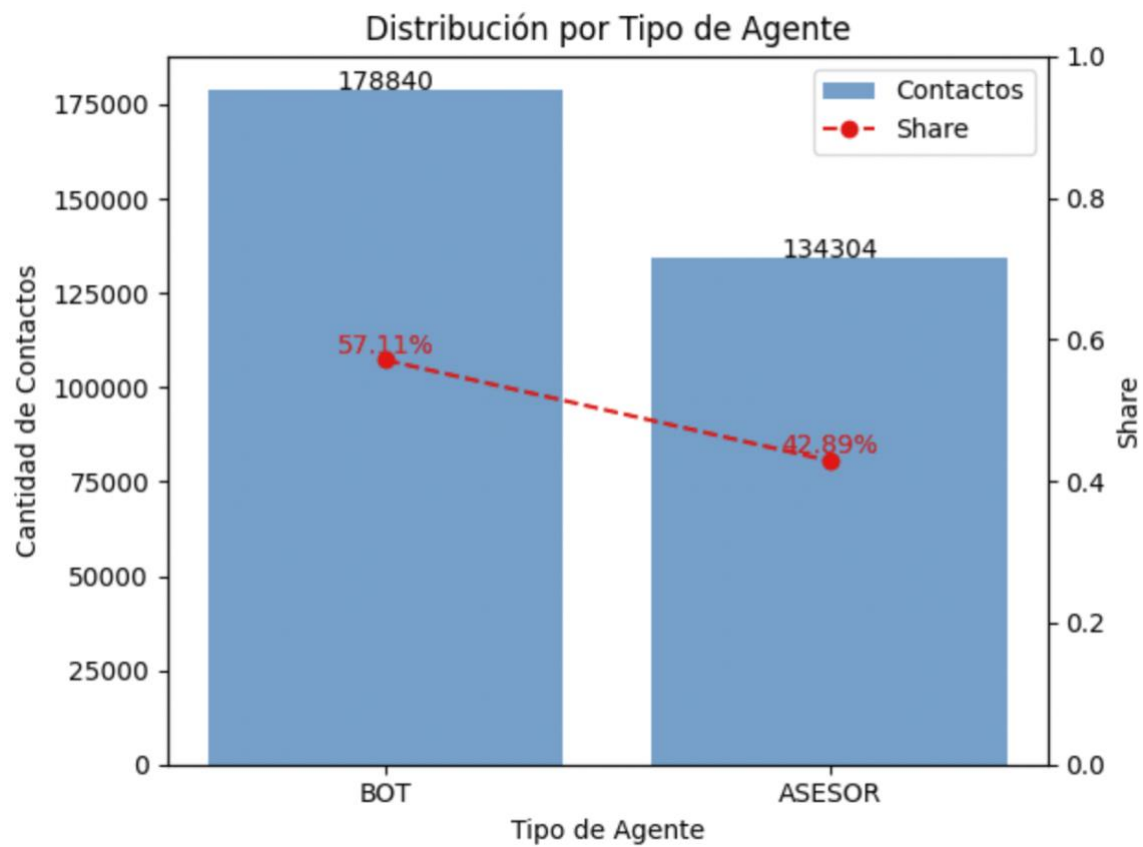


Ilustración 3. Distribución de interacciones por tipo de agente

Esta observación resalta la posibilidad de identificar oportunidades de mejora en este canal de atención.

La distribución por tipo de agente exhibe un comportamiento estable a nivel mensual:

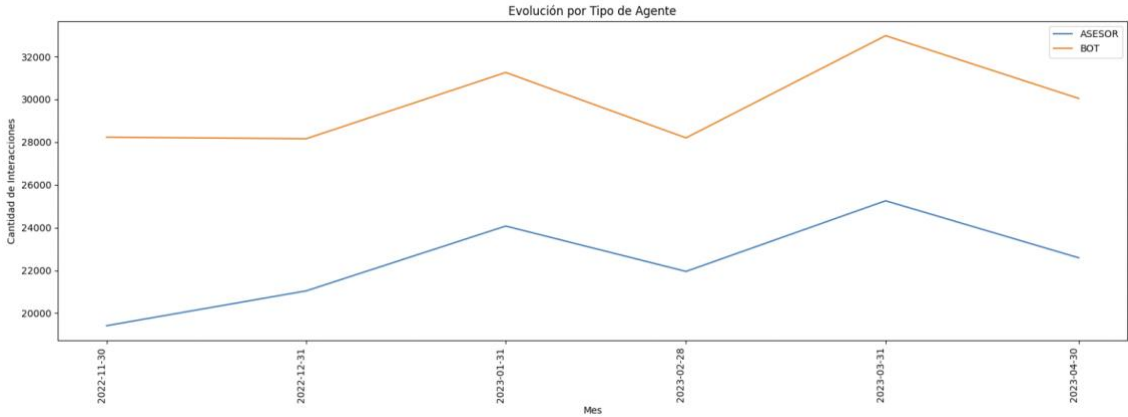


Ilustración 4. Evolución de interacciones por tipo de agente

Un asesor atiende aproximadamente 200 interacciones mensuales, dedicando en promedio 14 minutos por interacción. Mientras que las interacciones con el *chatbot* suelen durar 4 minutos en promedio.

En la mayor parte de las interacciones los clientes envían menos de 10 mensajes:

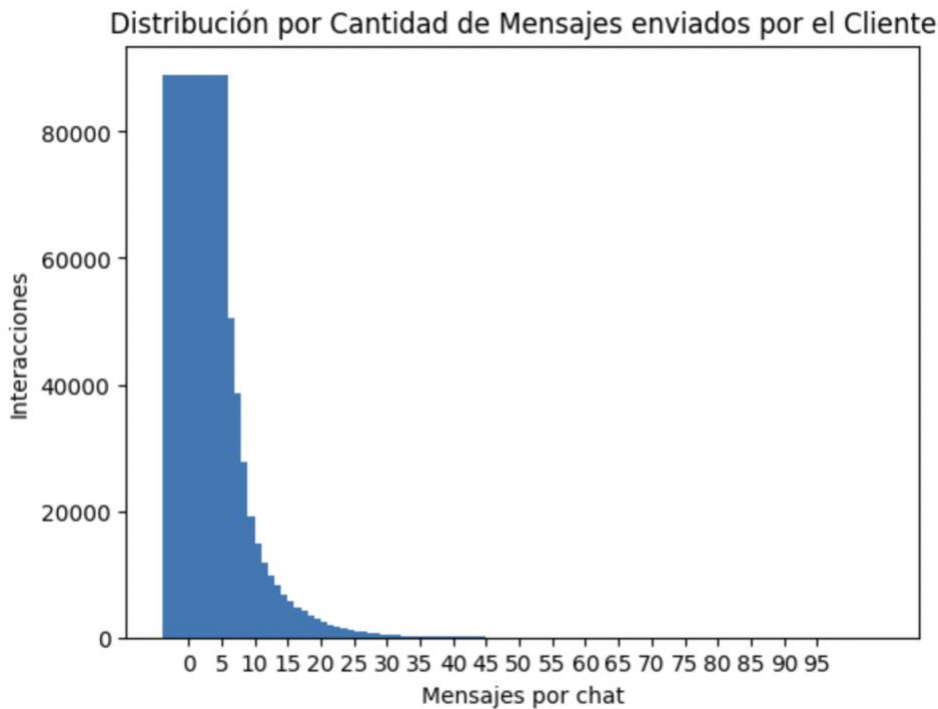


Ilustración 5. Cantidad de mensajes por interacción

Este comportamiento era previsible dado que en las interacciones a través de WhatsApp se espera un contacto rápido y ágil. El desafío está en lograr evidenciar el motivo o intención detrás de estas interacciones.

Se decidió utilizar únicamente los mensajes que fueron enviados por el cliente ya que las respuestas por parte de los asesores o *chatbot* introducen ruido que no aportan al modelo.

Las respuestas del *chatbot* varían entre un grupo de respuestas predefinidas y en ciertas ocasiones suelen ser muy genéricas, sin aportar información relevante a la consulta del cliente. En contraparte, las respuestas de los asesores suelen ser muy específicas, conteniendo información que solo tiene sentido para ese contexto delimitado.

A través de una nube de palabras se buscó identificar las palabras que aparecen con mayor frecuencia en los mensajes enviados por los clientes.

3.4.Lematización

Con el objetivo de reducir la dimensionalidad y mejorar la eficacia de los modelos se aplicó un proceso de lematización donde se llevó cada una de las palabras a su forma base (o lema). Esto permite que las palabras luego puedan ser analizadas y comparadas de manera más efectiva ya que permite que los modelos de NLP se enfoquen en la esencia del significado de las palabras, independientemente de las formas específicas en las que aparecen. Además, ayuda a reconocer y capturar relaciones semánticas entre palabras. Lematizar también contribuye a que los modelos generalicen mejor sobre datos nuevos y desconocidos ya que se centran en los aspectos más esenciales de las palabras y eliminan detalles específicos.

Para llevar a cabo la lematización de los textos se utilizó la librería Spacy que permite la carga de modelos preentrenados en grandes corpus de datos. En este caso, se utilizó el modelo preentrenado *es_dep_news_trf* basado en Transformers [1].

Una ventaja que ofrece este modelo es la capacidad de personalización mediante la adición de reglas y atributos personalizados. Esta funcionalidad resultó útil para unificar ciertas palabras que, en el contexto del negocio, comparten el mismo significado o lema. Por ejemplo: equipo, celular, aparato y cel, utilizados frecuentemente para referirse a un dispositivo móvil, o plan y contrato, usados de forma intercambiable para referirse a los planes de líneas móviles. También resultó útil para llevar a cabo correcciones gramaticales como por ejemplo corregir las formas conocidas de escribir incorrectamente roaming.

3.5. Verbos

Se identificó una gran presencia de verbos cotidianos (por ejemplo: querer, saber, etc.) que dificultaban el análisis ya que aparecían repetidamente a lo largo de las conversaciones. Con la asistencia del modelo preentrenado que también se utilizó para lematizar se logró identificar y eliminar las palabras verbales. Sin embargo, se observó que se estaban perdiendo algunas palabras de relevancia en el contexto de negocio (por ejemplo: recargar, renovar, baja, etc.). Para estas se añadieron reglas personalizadas que permitieran preservar palabras clave y de interés en el contexto.

3.6. DataFrame de entrenamiento.

A partir del análisis y limpieza de los datos, se obtuvo un dataframe con 227.495 registros para desarrollar y entrenar los modelos de NLP.

El mismo tiene la siguiente estructura:

	Id	client_msg	client_msg_trf
0	0008UaHRDM7P5EYS	Portabilidad a [REDACTED] quisiera saber cómo pa...	portabilidad plan numero
1	0008UaHRDM7P5F3S	Hola 6 Quiero darle la baja un contrato Buen d...	baja plan baje plan aca numero
2	0008UaHRDM7P5H1C	Demás me llega un mensaje o tu me avisas Como ...	demas mensaje habilitado gb
3	0008UaHRDM7P5H1M	Portabilidad a [REDACTED] Hola qué equipos tiene...	portabilidad equipo
4	0008UaHRDM7P5H2B	Hola buenos días! Celulares 2 Hola estoy busca...	equipo equipo almacenamiento bastante grande p...
5	0008UaHRDM7P5H2Q	Me interesa Si me interesa el nokia C20 2 Hola...	nokia c noelia equipo targeta compañera
6	0008UaHRDM7P5H3D	Buenass Tengo para renovar contrato Hola Yosel...	s renovar plan yoselin hermana deberia renovar...
7	0008UaHRDM7P5GXS	Hola Buenos días Yo quisiera sacar un aparato ...	equipo plan equipo plan
8	0008UaHRDM7P5H0H	Hola pase a plan cuenta controlada y los movim...	plan cuenta controlado movimiento acumulado movis
9	0008UaHRDM7P5H0K	Portabilidad a [REDACTED] Buenos días quería con...	portabilidad requisito equipo trabajador indep...

Ilustración 7. Estructura de DataFrame de entrenamiento

3.6.1. Descripción de las columnas

Id: Identificador interno de la interacción.

Client_msg: mensajes enviados por el cliente concatenados en un único string.

Client_msg_trf: mensajes enviados por el cliente luego de realizadas tareas de limpieza y concatenados en una única cadena de texto.

3.6.2. Descripción del contenido

Al analizar la nube de palabras que surge a partir de este nuevo dataframe se observa que algunas palabras previamente presentes se mantienen y se logran limpiar las que generan ruido. Al mismo tiempo se visualizan nuevas palabras que evidencian la presencia de nuevos tópicos. Entre ellas se encuentran palabras como roaming, portabilidad, saldo, paquete e incluso iPhone.

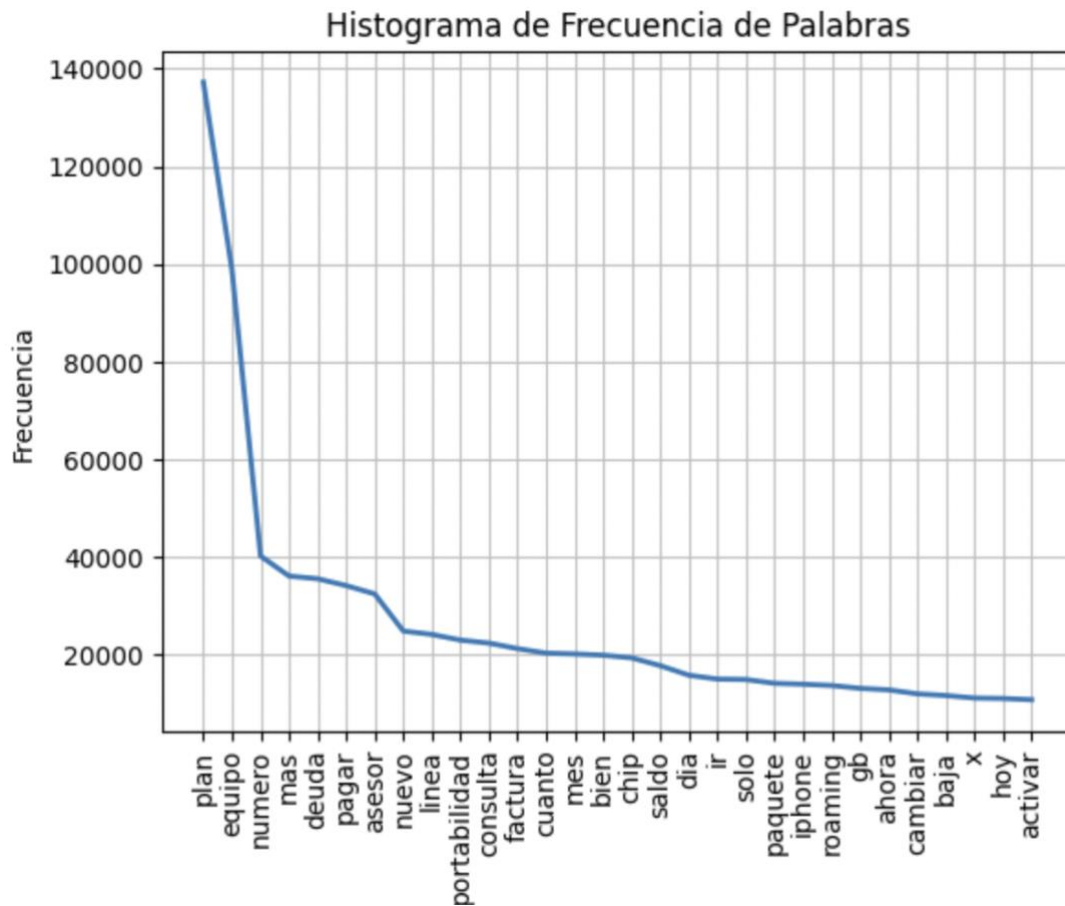


Ilustración 9. Histograma de frecuencia de palabras

Tal como evidencia el grafico de nube, las palabras plan y equipo son las que se destacan con mayor frecuencia en los mensajes. En particular la palabra plan se presenta con una frecuencia de aparición de 1 en cada 2 registros, lo que sugiere su alta relevancia en los mensajes enviados por los clientes. Cabe destacar que, en este contexto, la palabra plan puede hacer referencia al plan contratado por el cliente o a la obtención de nuevos planes. Será de interés poder diferenciarlos en la identificación de motivos.

4. Modelo

La implementación llevada a cabo es una clusterización de *embeddings* generados por un modelo pre entrenado basado en Transformers [1], más específicamente BERT (Bidirectional Encoder Representations from Transformers) [2].

4.1. Transformers

La arquitectura de Transformers [1] se encuentra en el estado del arte para las tareas de procesamiento de lenguaje, es reconocida por su capacidad para capturar relaciones semánticas complejas en el lenguaje. Es por esto, que en esta implementación es utilizada para generación de *embeddings*.

Transformers logra capturar relaciones semánticas complejas en el lenguaje gracias a su arquitectura basada en *Attention* y *Self Attention* y su entrenamiento con gran cantidad de texto.

Self Attention permite a cada palabra de una secuencia interactuar con todas las demás palabras de esa misma secuencia, ponderando la importancia relativa de cada palabra en función de su contexto. Esto permite que el modelo se centre en las palabras más relevantes dependiendo del contexto, lo que posibilita capturar relaciones semánticas complejas tanto a corto como a largo plazo. Para esto calcula ponderadores para cada palabra, llamados *Attention Weights*, que determinan cuánta atención se le debe prestar a cada palabra. Estos pesos se calculan comparando cada palabra con todas las demás palabras de la secuencia utilizando medidas, como por ejemplo, la similitud del coseno.

La arquitectura de Transformers tiene una arquitectura *Multi-Head Attention*, lo que significa que el modelo puede enfocarse en diferentes aspectos de las relaciones semánticas en paralelo capturando patrones complejos y dependencias del lenguaje. Cada cabeza de atención realiza sus cálculos de *Attention Weights* para cada palabra y luego los pesos de cada una de las salidas de las cabezas de atención se promedian para generar la representación final.

Durante el entrenamiento se alimenta al modelo con grandes cantidades de texto y se optimizan sus parámetros para predecir la siguiente palabra en una secuencia. De esta forma el modelo aprende a capturar relaciones semánticas y dependencias sintácticas en el lenguaje.

4.2.BERT

BERT [2] es un modelo de lenguaje pre entrenado basado en Transformers que tiene la capacidad de entender y generar texto en un contexto bidireccional. Es decir, puede procesar y comprender palabras y oraciones en relación con las palabras de su contexto anterior y posterior.

BERT utiliza la arquitectura de Transformers para modelar el contexto bidireccional. En lugar de utilizar una capa de atención unidireccional, como en los Transformers originales, BERT implementa una versión bidireccional, capturando el contexto de las palabras de manera más completa.

4.3.Embeddings

Los *embeddings* son una representación vectorial de un texto donde palabras con significados similares tienen vectores similares. El propósito principal es capturar características semánticas de las palabras (más que sintácticas o gramaticales) en función de su contexto en el texto. De esta manera las palabras son transformadas en valores numéricos que pueden ser utilizados por modelos de aprendizaje automático.

Para crear estos *embeddings* se utilizó Sentence-BERT (SBERT) [3], una variante del modelo pre entrenado BERT que se basa en estructuras de redes siamesas y triplets para capturar representaciones semánticas de las oraciones. Estas redes se utilizan para aprender y medir la similitud entre oraciones. Las redes siamesas comparan pares de texto, mientras que las redes triplets comparan un ancla, uno positivo (similar al ancla) y otro negativo (distinto del ancla) para entrenar al modelo. Mediante esta técnica se logra que las oraciones con significados similares estén representadas por vectores cercanos en el espacio, lo que permite luego poder compararlas utilizando la similitud del coseno como métrica de distancia.

La implementación de SBERT se llevó a cabo mediante la librería de Python *sentence-transformer* junto con el modelo pre entrenado *paraphrase-multilingual-mpnet-base-v2*[4]. Este modelo fue específicamente desarrollado y entrenado en la búsqueda semántica, siendo capaz de identificar textos relevantes a partir de una pregunta dada.

Los *embeddings* generados para el dataframe de entrenamiento fueron guardados en un archivo de tipo *joblib* que permitió poder reutilizarlos en distintas pruebas sin necesidad de regenerarlos en cada ejecución.

4.4. Clusterización

Los *embeddings* generados tienen una dimensionalidad de 768 características por lo que será necesario llevar a cabo una reducción de la dimensionalidad para lograr realizar una clasificación efectiva. Los modelos de clasificación que utilizan medidas de distancia como la Euclidiana presentan dificultades al manejar datos con alta dimensionalidad ya que la medida pierde significado al utilizarse en dimensiones tan altas.

4.4.1. UMAP

Para abordar esta tarea se recurrió al algoritmo UMAP (Uniform Manifold Approximation and Projection) [6]. UMAP ha demostrado ser altamente efectivo ya que permite conservar gran parte de la estructura de alta dimensionalidad en un espacio de menor dimensionalidad. Al utilizar UMAP se logra reducir la dimensionalidad de los *embeddings* sin perder información crucial para la clasificación.

4.4.2. HDBSCAN

Para la clasificación se utiliza HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise). HDBSCAN, es una técnica de clasificación que se basa en la densidad y jerarquía de los datos. Tiene la capacidad de encontrar clases de diferentes formas y tamaños, y es robusto frente a ruido y valores atípicos en los datos.

Una de las ventajas principales de este algoritmo es que no es necesario definir previamente la cantidad de clases. Esto resulta especialmente útil en este caso dado que no se cuenta con datos etiquetados ni se conoce la cantidad exacta de motivos a identificar en las interacciones.

4.5. Grid Search

Tanto UMAP como HDBSCAN tienen hiperparámetros configurables y la elección de estos puede tener un gran impacto en la eficacia de los modelos. Para encontrar la combinación óptima de los mismos se recurrió a la técnica de *Grid Search*. Esta técnica consiste en explorar las diferentes combinaciones de hiperparámetros definidos en una grilla de valores y para cada

una de ellas se entrena y se evalúa el modelo respecto al resultado obtenido en la métrica objetivo definida. En este caso se tomó como métrica la puntuación de *relative_validity_* que es una aproximación a Density Based Cluster Validity (DBCV) [7]. DBCV es una medida utilizada para evaluar algoritmos de clasificación basados en densidad que busca medir la densidad interna y la densidad externa de las clases para determinar qué tan bien definidos y separados están.

Los hiperparámetros incluidos en esta búsqueda fueron los siguientes:

Para UMAP:

n_neighbors: Especifica el número de vecinos utilizados para construir el grafo de vecinos. Controla la conectividad local en el espacio de alta dimensión. Valores más bajos capturan estructuras locales, mientras que valores más altos capturan estructuras globales. Por defecto 15.

n_components: Especifica la dimensionalidad de los *embeddings* luego de aplicar la reducción. Por defecto 2.

Para HDBSCAN:

min_cluster_size: Establece el tamaño mínimo requerido para formar un clúster. Los puntos que no cumplan con este tamaño mínimo se consideran ruido o *outliers*. Por defecto 5.

Para definir los valores de hiperparámetros sobre los que iterar, se tuvieron en cuenta los valores por defecto de los algoritmos y la dimensionalidad y cantidad de datos de entrenamiento.

En el Anexo 4, se puede observar la grilla de hiperparámetros iterada, con los resultados para cada combinación. La misma se encuentra ordenada descendientemente por el resultado de la métrica definida.

Los hiperparámetros elegidos para llevar a cabo la ejecución de UMAP fueron 10 para *n_neighbors* y de 20 para *n_components*. Asimismo, en el caso de HDBSCAN, se estableció 300 para *min_size_cluster*.

Estos hiperparámetros fueron los que obtuvieron mejores resultados en *relative_validity_*, pero también se tuvieron en cuenta otras particularidades. Era esencial generar un número de clases significativo sin que alguno de ellos predominara demasiado ya que podría llevar a la falta de

identificación de motivos específicos. A partir del análisis de wordclouds, se podían identificar al menos 10 amplios tópicos y dado que el objetivo es encontrar motivos que el *chatbot* aún no sabe resolver, se optó por una implementación que permitiera clasificar al menos 35 tópicos, pero no más de 100. De ser necesario, se puede efectuar una reducción de tópicos posterior.

4.6. Caracterización

Luego de obtener las clases, es necesario realizar una caracterización de estos, para eso se implementa una solución de TF-IDF (Term Frequency-Inverse Document Frequency) basada en clases. De esta forma se logran extraer las características que hacen única a cada clase.

4.6.1. C-TF-IDF

Grootendorst introduce una variante de TF-IDF (C-TF-IDF) [8] que al utilizar TF-IDF en el contexto de cada clase se prioriza la importancia de las palabras o términos que son específicos y destacados dentro de ese grupo en particular. De esta forma se logran identificar las palabras o términos clave que contribuyen significativamente a la representación del motivo de contacto de cada clase de interacciones.

En este proceso la frecuencia de cada palabra t se extrae para cada clase i y luego se divide por el total de palabras w . Esto actúa como una forma de regularización para las palabras que son recurrentes dentro de una clase específica. Posteriormente la cantidad total de interacciones m se divide por la frecuencia de la palabra t a lo largo de todas las clases n .

$$c - TF - IDF_i = \frac{t_i}{w_i} \times \log \frac{m}{\sum_j^n t_j}$$

Ilustración 10. Fórmula de cálculo para c-TF-IDF

Utilizando esta función se obtienen las 20 palabras con los puntajes más altos de c-TF-IDF. Estos puntajes reflejan la importancia de las palabras en relación con el contenido y el contexto de cada clúster. Cuanto mayor sea el puntaje de c-TF-IDF de una palabra, mayor será su representatividad del motivo en esa clase en particular.

4.6.2. Auto etiquetado

Para poder automatizar este proceso y que sea de utilidad para ejecuciones futuras del modelo se desarrolló una función inspirada en el paper de Liu et al. [9] donde se asigna una etiqueta descriptiva a cada clase a partir del par de acción-objeto más frecuente.

Dado el complejo contexto de negocio en este caso se decidió preservar las palabras con mayor puntaje obtenido en C-TF-IDF, seleccionando para cada una de las categorías gramaticales principales (verbos, sustantivos, pronombres y adjetivos) las dos palabras con mayor puntaje, siempre y cuando dicho puntaje fuera mayor a un umbral de relevancia predefinido.

Para llevar a cabo esta tarea se utilizó la librería Spacy anteriormente mencionada que ofrece la clasificación gramatical para cada palabra. Este enfoque permite obtener una representación descriptiva de los motivos de contacto presentes en cada grupo de interacciones de forma automatizada.

4.7.Reducción de Tópicos

Dado que en la etapa de clasificación se obtuvieron un total de 50 clases distintas, es necesario recurrir a una técnica que facilite la reducción de esta cantidad para que los resultados sean comprensibles y manejables.

Para lograr consolidar clases similares entre sí y obtener clases más amplias y representativas se utilizó una técnica que comparando los resultados de c-TF-IDF entre clases agrupa las más similares utilizando la métrica de similitud del coseno. Para esto, se definió un umbral de similitud a partir del cual se considera conveniente reducir, en este caso, 0.75 y se procede a consolidarlos. En el Anexo 5 se puede visualizar la matriz de similitud entre las clases. Luego de cada consolidación es necesario recalcular los puntajes de c-TF-IDF para actualizar la representación de las clases resultantes y poder repetir el procedimiento.

4.8. Resultados

Finalmente, se obtuvieron 34 clases con la siguiente distribución:

label	size
otros	114499
deuda_pagar_asesor	19465
portabilidad_samsung_equipo_nuevo_pagar_galaxy	19002
chip_activar_numero_asesor_nuevo_pin	11134
iphone_pro_equipo_max_nuevo_comprar	9242
roaming_activar_argentina_brasil_activado_uruguay	8806
llamar_asesor_llamada	5258
equipo_plan_cedula_incluido	5074
nuevo_equipo_plan_pagar_cedula	4696
paquete_gigas_saldo_cargar_app_solo	3090
tarjeta_credito_pagar_recarga_pago_solo	2536
asesor	2178
app_codigo_asesor_recarga_solo_cargar	2075
plan	1899
plan_terminar_baja_numero_solo	1628
wifi_modem_dato_señal_solo	1618
saldo	1561
whatsapp_gratis_solo_numero_llamada_llamar	1545
portabilidad	1308
informacion_equipo_general_info_movil	1308
telefono_plan_nuevo_comprar_gratis_phone	1199
consulta_consulto	1051
factura_informacion_general_pagar_info	1012
paquete_gigante_comprar_excedente_excedent_duo	930
internet_paquete_dato_red_cargar	802
pedido_dueda_solicitud_ci	798
menu_deuda_saldo_principal_internet_cambiar	671
renovar_plan_asesor_baja_perfecto	569
motorola_moto_equipo_nuevo_plan_pagar	565
factura_fatura_adjunto_consuta	539
favorito_numero_cambiar_llamada	420
pago_consulta_cosnulta	377
recarga_cuba_prestmo_acteditariar_dacteo	327
activar_activado_activacion_linea_esim_sim	313

Tabla 1. Distribución por clústers

Como se había evidenciado durante la búsqueda de hiperparámetros, la clase *otros* es la más predominante ya que corresponde a los *outliers* que el modelo no logró clasificar. Esto se explica por la presencia de interacciones que no hacen referencia a nada en particular o que tocan demasiados temas a la vez.

5. Evaluación del Modelo

Dado que no se cuentan con datos etiquetados se recurrió al etiquetado manual de expertos del dominio. Se solicitó al equipo de Business Intelligence de la organización que colaborara con el etiquetado de una muestra de datos que luego sirviera para evaluar los resultados del modelo. Los integrantes de este equipo están familiarizados tanto con los datos como con el negocio dado que en diversas ocasiones se han enfrentado a analizar información de esta fuente o similares.

Para llevar a cabo esta tarea se les disponibilizó una interfaz web basada en Gradio donde aleatoriamente se mostraba una interacción y se les solicitaba seleccionar la etiqueta correspondiente. Como posibles etiquetas, se listaban los motivos identificados por el modelo. Se pueden observar las pantallas de esta interfaz en el Anexo 6.

Una vez obtenida la muestra etiquetada se procedió a comparar con las etiquetas asignadas por el modelo para poder evaluar su performance.

	precision	recall	f1-score	support
app_codigo_asesor_recarga_solo_cargar	1	0,5	0,67	2
asesor	1	0,57	0,73	7
chip_activar_numero_asesor_nuevo_pin	0,55	1	0,71	6
consulta_consulta	1	1	1	1
deuda_pagar_asesor	1	0,55	0,71	42
equipo_plan_cedula_incluido	0,25	0,33	0,29	3
factura_informacion_general_pagar_info	1	0,08	0,14	13
favorito_numero_cambiar_llamada	1	0,2	0,33	5
informacion_equipo_general_info_movil	1	0,11	0,2	18
internet_paquete_dato_red_cargar	1	0,5	0,67	2
iphone_pro_equipo_max_nuevo_comprar	0,92	0,79	0,85	14
llamar_asesor_llamada	0,75	1	0,86	3
nuevo_equipo_plan_pagar_cedula	0,6	0,17	0,26	18
otros	0,29	0,95	0,44	44
paquete_gigante_comprar_excedente_excedent_duo	1	1	1	1
paquete_gigas_saldo_cargar_app_solo	0,33	0,14	0,2	7
pedido_dueda_solicitud_ci	0	0	0	6
plan	1	0,14	0,25	7
plan_terminar_baja_numero_solo	0	0	0	8
portabilidad	1	0,5	0,67	4
portabilidad_samsung_equipo_nuevo_pagar_galaxy	0,54	0,88	0,67	8
renovar_plan_asesor_baja_perfecto	1	0,25	0,4	4
roaming_activar_argentina_brasil_activado_uruguay	1	0,85	0,92	13
saldo	1	0,44	0,62	9
tarjeta_credito_pagar_recarga_pago_solo	0	0	0	5
telefono_plan_nuevo_comprar_gratis_phone	1	1	1	1
whatsapp_gratis_solo_numero_llamada_llamar	0,67	1	0,8	2
wifi_modem_dato_señal_solo	0	0	0	8
accuracy			0,5	261
macro avg	0,71	0,5	0,51	261
weighted avg	0,69	0,5	0,47	261

Tabla 2. Reporte de evaluación del modelo

Lo primero que se puede observar es que no todas las clases están presentes en el análisis de resultados. Esto se debe a que son las clases con menor cantidad de registros, y dado que se obtuvo una muestra aleatoria para etiquetar, estas no quedaron representadas en la muestra.

Se puede observar que algunas clases muestran una precisión del 100% o valores cercanos, lo que indica que el modelo coincide en gran medida con las etiquetas asignadas por los usuarios. Sin embargo, también existen clases con una precisión muy baja, incluso llegando a 0. Por ejemplo, las clases `plan_terminar_baja_numero_solo` o `wifi_modem_dato_señal_solo`. La primera se refiere a clientes que desean dar de baja sus servicios, y la segunda a clientes que experimentan problemas con él, ambas son clases sensibles en las que sería deseable tener un mejor rendimiento.

Clases como `factura_informacion_general_pagar_info` o `favorito_numero_cambiar_llamada` lograron un buen rendimiento en precisión. Sin embargo su resultado en recall no es tan alto, indicando que el modelo no logra detectar correctamente todos los registros para estas clases.

Considerando la cantidad de clases presentes y aunque es deseable continuar refinado y ajustando el modelo, el valor obtenido en *weighted avg precision* es muy bueno para el propósito de este proyecto. Dado que el objetivo principal es la identificación de los motivos de contacto más relevantes y las clases con mayor cantidad de interacciones tienen un alto nivel de precisión, el modelo proporciona una base sólida para tomar acciones y definir mejoras en los procesos de negocio.

6. Conclusiones y próximos pasos

Conocer los motivos de contacto de los clientes ofrece oportunidades de mejora en varios frentes, ya sea para mejorar los productos y servicios, automatizar procesos o mejorar la satisfacción de los clientes, entre otros.

Al analizar los motivos identificados por el modelo se observa que, como era de esperar, muchos de ellos ya han sido resueltos a través de los casos de uso del *chatbot*. Sin embargo, se detectaron algunos que ofrecen oportunidades de mejora.

Con la clase `roaming_activar_argentina_brasil_activado_uruguay` se evidencian consultas recurrentes sobre la activación de roaming. Al simular una conversación vía WhatsApp con el *chatbot* este no logra resolver el inconveniente, sino que deriva el cliente a otro canal o con un asesor. Se puede visualizar dicha interacción en el Anexo 7.

Sería interesante poder implementar una solución automática por esta vía, es decir, poder validar el estado del servicio en el sistema y realizar la activación de ser necesario.

Se calculó el tiempo que los asesores dedican a estas interacciones y, en promedio, se destinan 4 horas al día. Esto equivale a la mitad del tiempo disponible de un asesor. Se debe tener en cuenta que dentro de las interacciones que no fueron derivadas a un asesor pueden existir clientes insatisfechos que se verían beneficiados de implementar la solución propuesta.

De realizarse la automatización mencionada previamente, la eficiencia generada brinda la oportunidad de dedicar tiempo a abordar otro de los motivos identificados, la clase `llamar_asesor_llamada` representa a aquellos clientes que desean ser contactados por un asesor, no son clientes que necesitan el contacto inmediato, sino que pueden esperar a ser llamados cuando un asesor quede disponible. Aunque esta sugerencia no necesariamente se traduce en ganancias en términos de eficiencia, es una decisión que puede contribuir significativamente a la satisfacción del cliente con la organización.

Para las clases que hacen referencia a la adquisición de productos o servicios (planes, portabilidad, equipos, Iphone), se puede analizar su conversión posterior. Es decir, si realmente se efectuó la compra o contratación. Esta información luego se puede utilizar para alimentar las campañas de venta de telemarketing, asignándoles mayor prioridad a los clientes que demostraron interés sobre ciertos productos y servicios.

Las sugerencias mencionadas son ejemplos de cómo se pueden aplicar los resultados del modelo. Sin embargo, es importante destacar que los hallazgos de este modelo pueden generar múltiples beneficios, dependiendo del área usuaria de la organización que los utilice. Por esta razón, se recomienda continuar refinando el modelo con el objetivo de obtener mejores resultados, especialmente en términos de *recall* para asegurar que el modelo pueda identificar cualquier motivo relevante que hasta el momento haya pasado desapercibido. Además, se recomienda ejecutar el modelo periódicamente (semestral o trimestralmente) para identificar posibles cambios en el contexto del negocio que puedan generar nuevas clases o modificar la relevancia de las existentes.

7. Referencias Bibliográficas

- [1] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin, I. 2017. **Attention Is All You Need** [Online] Disponible en Internet: [arXiv:1706.03762](https://arxiv.org/abs/1706.03762)
- [2] Devlin J., Chang M., Lee K., Toutanova K.. 2019. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. [Online] Disponible en Internet: [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
- [3] Reimers N., Gurevych I., 2020. **Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation** [Online] Disponible en Internet: [arXiv:2004.09813](https://arxiv.org/abs/2004.09813)
- [4] Reimers N., Gurevych I., 2019. **Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks** [Online] Disponible en Internet: [arXiv:1908.10084](https://arxiv.org/abs/1908.10084)
- [5] Aggarwal C., Hinneburg A., Keim D., 2001. **On the Surprising Behavior of Distance Metrics in High Dimensional Space** [Online] Disponible en Internet: doi.org/10.1007/3-540-44503-X_27
- [6] McInnes L., Healy J., Melville J., 2018. **UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction** [Online] Disponible en Internet: doi.org/10.21105/joss.00861
- [7] Moulavi D., Jaskowiak P.A., Campello R.J., Zimek, A., Sander, J., 2014. **Density-Based Clustering Validation**. [Online] Disponible en Internet: [10.1137/1.9781611973440.96](https://doi.org/10.1137/1.9781611973440.96)
- [8] Grootendorst M., 2020. **At the Intersection of AI and Psychology Creating a class-based TF-IDF with Scikit-Learn** [Online] Disponible en Internet: maartengrootendorst.com/blog/ctfidf/
- [9] Pengfei L., Youzhang N., King Keung W., Kun L., Meng H., 2021. **Open Intent Discovery through Unsupervised Semantic Clustering and Dependency Parsing** [Online] Disponible en Internet: [arXiv:2104.12114v2](https://arxiv.org/abs/2104.12114v2)

8. Anexos

8.1. Anexo 1: Casos de uso automáticos implementados.

Son casos de uso donde el *chatbot* se comunica con el sistema comercial para obtener información relativa a los productos y servicios del cliente.

- Consulta de Deuda
- Consulta de Pago
- Consulta de Excedentes
- Detalle de plan
- Estado de Portabilidad
- ¿Qué pasó con mi pedido?
- ¿Cuándo llega mi pedido?
- ¿Qué compré?
- Ver mis últimos pedidos
- Consulta de Ajustes
- Deuda SOS

8.2. Anexo 2: Ejemplo de Interacción

Ejemplo de cómo se almacena una interacción en formato XML:

```
<?xml version="1.0" ?>
<chatTranscript startAt="2022-10-27T04:22:51Z" sessionId="0008KaHK7GW592QC">
  <newParty userId="00FB635A079B7129" timeShift="0" visibility="ALL" eventId="1">
    <userInfo personId="" userNick="WhatsApp" userType="CLIENT" protocolType="FLEX" timeZoneOffset="-180" clientVersion="106"/>
    <userData>
      <item key="FirstName" type="string">WhatsApp</item>
      <item key="IdentifyOrCreateContact" type="string">3</item>
      <item key="LastName" type="string"></item>
      <item key="MediaType" type="string">chat</item>
      <item key="PhoneNumber" type="string"></item>
      <item key="TimeZone" type="string">-180</item>
    </userData>
    <configuration>
      <inactivityControl>
        <item key="message-alert" type="string">¡Hola! estamos disponibles para ayudarte con tu gestión y comentarte todos los beneficios que tenemos para vos.</item>
        <item key="message-close" type="string">Gracias por comunicarte.</item>
        <item key="timeout-alert" type="string">360</item>
        <item key="timeout-close" type="string">720</item>
      </inactivityControl>
    </configuration>
  </newParty>
  <newParty userId="00FB635A07A3712B" timeShift="8" visibility="ALL" eventId="2">
    <userInfo personId="" userNick="SofaBotTMK" userType="AGENT" protocolType="BASIC" timeZoneOffset="-1080000" clientVersion="106"/>
    <userData>
      <item key="GCTL_Chat_PartyStyle" type="string">BOT</item>
      <item key="GCTL_Chat_SetPartyStyle" type="string">BOT</item>
    </userData>
    <eventAttributes>
      <item key="GCTL_SYSTEM" type="kvlist">
        <item key="party-info" type="kvlist">
          <item key="style" type="string">BOT</item>
        </item>
      </item>
      <item key="chat-bot-startup-info" type="kvlist">
        <item key="ChatBotHoldup" type="string">true</item>
        <item key="ChatBotD" type="string">SofaBotTMK</item>
        <item key="ChatBotName" type="string">SofaBotTMK</item>
        <item key="StopBotOnAgentArrival" type="string">true</item>
        <item key="StopBotOnCustomerLeft" type="string">true</item>
        <item key="cbs" type="string">04CHAW33AY66J3GS</item>
        <item key="launch-mode" type="string">STARTED_BY_ESP</item>
      </item>
    </eventAttributes>
  </newParty>
  <message userId="00FB635A079B7129" timeShift="11" visibility="ALL" eventId="3">
    <msgText msgType="text">Quiero un nuevo equipo mas el plan</msgText>
  </message>
  <message userId="00FB635A07A3712B" timeShift="12" visibility="ALL" eventId="4">
    <msgText msgType="text"><span>¡Genial! Celular nuevo y un plan con muchos GIGAS para navegar hasta el cansancio 🤩</span>
    Mientras te atiende uno de nuestros asesores para ayudarte con la compra, te invitamos a ir viendo estas ofertas de celulares con plan que te pueden interesar: <a href="https://ofertas.com.uy/celulares/contratos/msgText">https://ofertas.com.uy/celulares/contratos/msgText</a>
  </message>
  <partyLeft userId="00FB635A07A3712B" timeShift="12" visibility="ALL" eventId="5" askerId="00FB635A07A3712B">
    <reason>left</reason>
  </partyLeft>
  <newParty userId="00FB635A07B8712C" timeShift="29" visibility="ALL" eventId="6">
    <userInfo personId="" userNick="" userType="EXTERNAL" protocolType="ESP" timeZoneOffset="0"/>
  </newParty>
  <message userId="00FB635A07B8712C" timeShift="29" visibility="ALL" eventId="7">
    <msgText>¡Hola! Te damos la bienvenida al Canal de ventas de . Nuestro horario de atención es de lunes a sábado de 9 a 21h. Conocé nuestras ofertas y dejanos tus datos que con gusto nos pondremos en cont
    acto contigo - bit.ly/</msgText>
  </message>
  <partyLeft userId="00FB635A079B7129" timeShift="34" visibility="ALL" eventId="8">
    <reason code="5">removed by the server</reason>
  </partyLeft>
</chatTranscript>
```

Ilustración 11. Ejemplo XML

8.3. Anexo 3: Ejemplo Click To Chat

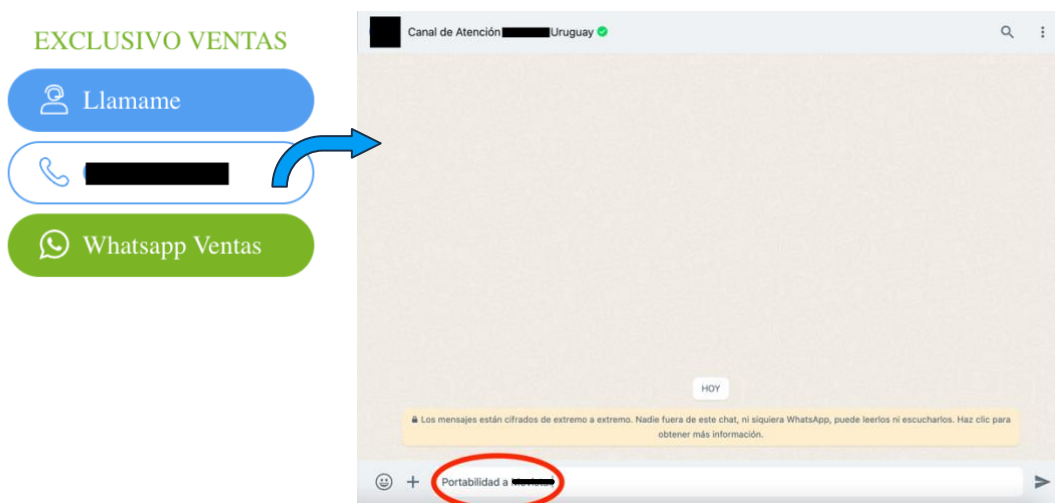


Ilustración 12. Ejemplo click to chat

8.4. Anexo 4: Resultados Grid Search

n_neighbors	n_components	min_cluster_size	label_count	score	coverage	max_cluster	params
10	20	300	50	0,40368	0,607688	11,1	(10, 20, 300)
20	20	200	77	0,37853	0,607855	8,3	(20, 20, 200)
15	5	100	168	0,36262	0,642023	8,5	(15, 5, 100)
15	5	200	83	0,33728	0,616123	8,5	(15, 5, 200)
10	5	100	174	0,33492	0,641372	8,2	(10, 5, 100)
15	15	100	175	0,33171	0,610321	8,5	(15, 15, 100)
10	10	100	167	0,33094	0,628005	8,2	(10, 10, 100)
20	10	200	77	0,2949	0,597055	8,5	(20, 10, 200)
20	20	100	160	0,28142	0,65408	10,3	(20, 20, 100)
20	5	100	173	0,28057	0,630801	8,2	(20, 5, 100)
15	10	100	171	0,27891	0,621671	8,2	(15, 10, 100)
15	20	200	73	0,27482	0,613446	8,4	(15, 20, 200)
20	15	300	32	0,27313	0,869021	50,4	(20, 15, 300)
10	20	100	186	0,26492	0,585855	8,2	(10, 20, 100)
20	10	100	175	0,26023	0,597996	8,1	(20, 10, 100)
20	5	200	78	0,25706	0,620554	8,3	(20, 5, 200)
15	20	100	173	0,2568	0,603196	8,4	(15, 20, 100)
20	5	400	27	0,24627	0,888613	53,8	(20, 5, 400)
20	15	100	167	0,23505	0,639838	8,2	(20, 15, 100)
10	15	100	183	0,23104	0,597099	8,2	(10, 15, 100)
10	20	400	26	0,23053	0,821926	47,2	(10, 20, 400)
10	10	200	50	0,23014	0,903572	52,2	(10, 10, 200)
15	10	200	48	0,22719	0,896609	51,5	(15, 10, 200)
20	20	300	31	0,22163	0,861755	50	(20, 20, 300)
20	20	400	25	0,21918	0,805934	44,9	(20, 20, 400)
20	5	300	34	0,21524	0,902068	53,7	(20, 5, 300)
15	20	300	50	0,20117	0,552434	8,4	(15, 20, 300)
20	15	200	49	0,19933	0,888877	50,5	(20, 15, 200)
10	15	200	47	0,19416	0,902723	52	(10, 15, 200)
10	5	400	26	0,19348	0,860797	50,7	(10, 5, 400)
15	10	400	24	0,18599	0,87052	51,6	(15, 10, 400)
20	15	400	24	0,18419	0,850792	49,8	(20, 15, 400)
15	10	300	31	0,18327	0,888683	52,5	(15, 10, 300)
15	5	300	36	0,18197	0,895549	52,7	(15, 5, 300)
20	20	500	24	0,17073	0,861465	51,1	(20, 20, 500)
10	15	300	30	0,17027	0,886595	52,3	(10, 15, 300)
15	5	400	27	0,16992	0,866731	50,9	(15, 5, 400)
10	5	200	52	0,16825	0,898354	51,5	(10, 5, 200)
15	10	500	23	0,16483	0,860911	51	(15, 10, 500)
15	5	500	22	0,15412	0,850225	50,2	(15, 5, 500)
10	5	500	24	0,15265	0,920253	59,6	(10, 5, 500)
10	20	200	41	0,14892	0,936755	61,6	(10, 20, 200)
10	10	500	22	0,14703	0,895136	54,4	(10, 10, 500)
10	10	300	35	0,14701	0,882586	51,4	(10, 10, 300)
10	5	300	33	0,14582	0,914578	54,8	(10, 5, 300)
10	15	400	24	0,14575	0,879061	52,5	(10, 15, 400)
20	10	400	24	0,14453	0,866274	51,5	(20, 10, 400)
15	15	300	31	0,14192	0,835583	46,3	(15, 15, 300)
10	10	400	25	0,13333	0,878978	52,5	(10, 10, 400)
20	5	500	23	0,13319	0,88598	53,8	(20, 5, 500)
10	20	500	21	0,13307	0,848005	50,5	(10, 20, 500)
15	15	200	42	0,13265	0,945863	61,7	(15, 15, 200)
15	15	400	24	0,12868	0,854085	49,9	(15, 15, 400)
10	15	500	22	0,12736	0,89569	54,6	(10, 15, 500)
20	15	500	24	0,12013	0,870951	51,6	(20, 15, 500)
20	10	300	30	0,11019	0,893136	53	(20, 10, 300)
20	10	500	23	0,10765	0,880591	52,7	(20, 10, 500)
15	20	500	24	0,09991	0,867966	51	(15, 20, 500)
15	15	500	23	0,09532	0,843838	49	(15, 15, 500)
15	20	400	25	0,09257	0,871936	51,5	(15, 20, 400)

Ilustración 13. Detalle de resultados de Grid Search

8.6. Anexo 6: Interfaz de etiquetado

Texto

asesor 094543148 Buenos días!!ayer hablé con un asesor para que me ayudará con un canje de movis por un paquete que nunca llegó y todavía nada era un paquete de 5gb por un día me descontó los movis pero el paquete nunca llegó no no tengo porque me los descontaron ayer por ese paquete ahi está ahí te mande captura el canje es de ayer no tengo otro numero

Etiqueta

activar_activado_activacion_linea_esim_sim app_codigo_aplicacion_asesor_solo_cargar asesor chip_activar_numero_asesor_nuevo_pin consulta_consulto

deuda_pagar_asesor equipo_plan_cedula_incluido factura_fatura_adjunto_consulta factura_informacion_general_pagar_info favorito_numero_cambiar_llamada

informacion_equipo_general_info_movil internet_paquete_dato_red_cargar iphone_pro_equipo_max_nuevo_comprar llamar_asesor_llamada

menu_deuda_saldo_principal_internet_cambiar motorola_moto_equipo_nuevo_plan_pagar nuevo_equipo_plan_pagar_cedula otros pago_consulta_cosulta

paquete_gigante_comprar_excedente_excedent_duo paquete_gigas_saldo_cargar_app_solo pedido_dueda_solicitud_ci plan plan_terminar_baja_numero_solo

portabilidad portabilidad_samsung_equipo_nuevo_pagar_galaxy recarga_cuba_prestmo_acteditariar_dacteo renovar_plan_asesor_baja_perfecto

roaming_activar_argentina_brasil_activado_uruguay saldo tarjeta_credito_pagar_recarga_pago_solo telefono_plan_nuevo_comprar_gratis_phone

whatsapp_gratis_solo_numero_llamada_llamar wifi_modem_dato_señal_solo

Etiquetar

Resultado

Texto: Hola buen dia Una pregunta se puede cambiar de compania y conservar el mismo número? Se puede cambiar de compania y usando el mismo número? Si Tengo q ir al local en Carmelo No uso contrato en la otra compania No me puede llamar q es más fácil y veo los contratos gracias Gustavo Silva 38418522.. Ancel Ese mismo Gracias

Etiqueta: portabilidad



Use via API  · Built with Gradio 

Ilustración 15. Captura de interfaz gráfica de etiquetado

8.7. Anexo 7: Interacción con chatbot sobre activación de roaming

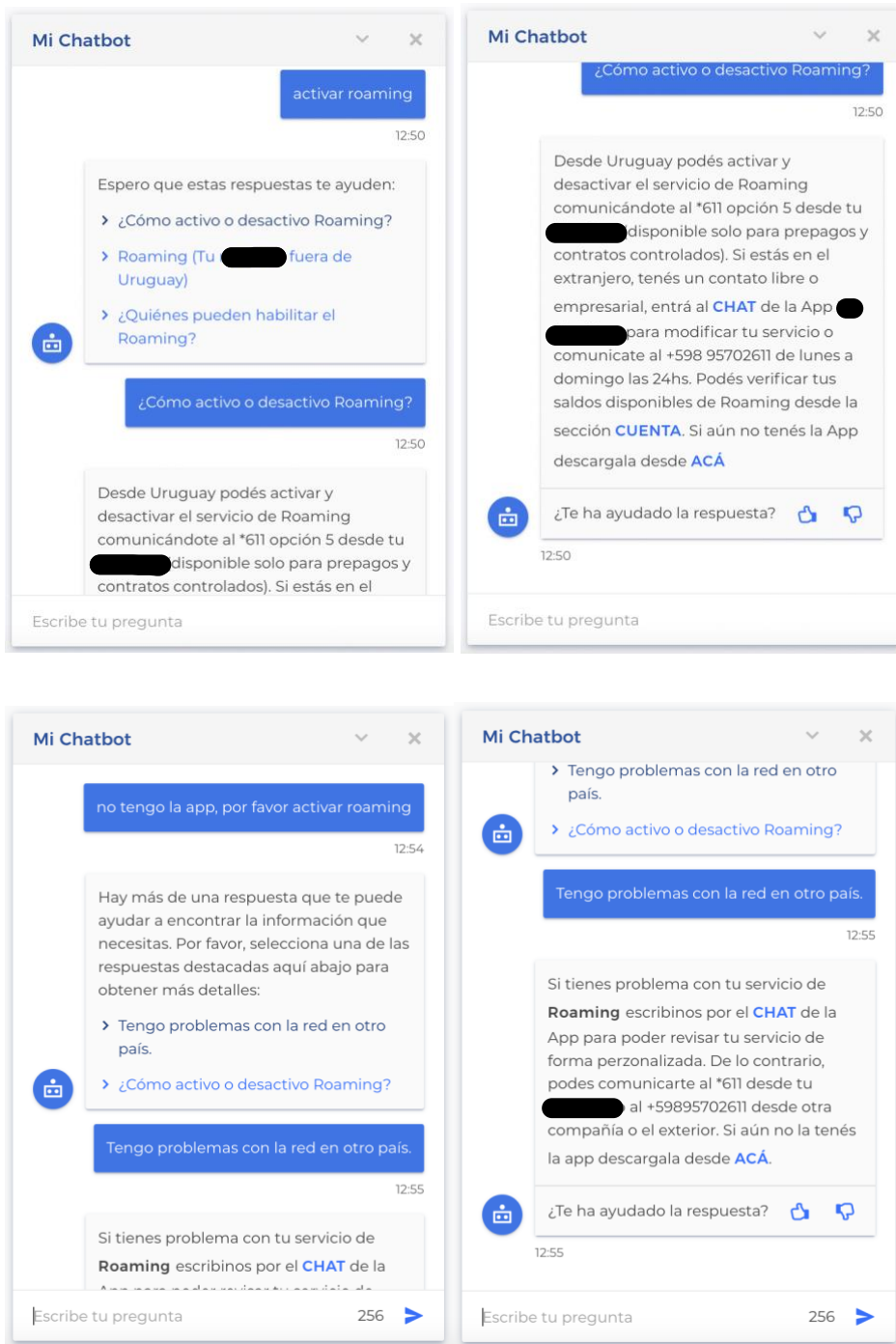


Ilustración 16. Interacción con chatbot sobre roaming